

**IMPLEMENTASI TEKNIK WEIGHTED K NEAREST NEIGHBOR UNTUK
MEMPERBAIKI DATA SET YANG HILANG PADA PENYAKIT KANKER
PAYUDARA**

Watsiqatut Dianah

(150411100124)

Dosen Pembimbing 1 : Mula'ab S.Si., M., Kom

Dosen Pembimbing 2 : Dr. Yeni Kustiyahningsih, S. Kom., M. Kom

ABSTRAK

Data set yang hilang bukan merupakan masalah yang utama untuk studi yang bersifat substansif pada umumnya, tapi akan mengakibatkan masalah yang serius jika tidak ditangani secara tepat. Banyak kasus yang ditemukan pada *database* yang datanya tidak lengkap yang setengahnya hilang dari data set, sehingga sulit sekali untuk melakukan penambangan data, dimana hanya data yang lengkap untuk dilakukan analisis. Dengan demikian, dibutuhkan sistem yang lebih praktis dan mempunyai kemampuan dalam mengisi nilai yang hilang dengan nilai – nilai yang mungkin berdasarkan informasi yang tersedia pada data set tersebut. Terdapat banyak metode untuk mengimputasi data hilang. Metode *weighted k-Nearest Neighbor* (wkNN) merupakan pengembangan dari metode *k-Nearest Neighbor* (kNN) dan memiliki prinsip yang sama dengan metode *k-Nearest Neighbor* (kNN) yaitu mencari jarak terdekat antara data yang akan diuji dengan sejumlah k tetangga terdekatnya dalam data latih untuk menentukan nilai bobot. Pada penelitian ini dataset diambil dari jumlah banyak data kanker payudara (*Breast Cancer Wiscosin*) diperoleh dari University of Wisconsin Hospitals, Madison dari Dr. William H. Wolberg. Berdasarkan hasil uji coba sebanyak 683 data, 9 atribut, 5 data yang dihilangkan secara berurutan dan sembarang/acak. Hasil yang didapat menggunakan metode *weighted k-Nearest Neighbor* (wkNN) ialah menggunakan k tetangga terdekat yaitu k (3,5,7,9,11) dan hasil yang terbaik ialah terdapat pada $k = 11$, karena semakin besar k yang digunakan maka semakin baik nilai SSE yang diperoleh.

Kata kunci : data hilang, wkNN, tetangga terdekat, SSE

IMPLEMENTATION OF WEIGHTED K NEAREST NEIGHBOR TECHNIQUE FOR IMPROVING THE MISSING DATA SET ON CANCER DISEASE BREAST

Watsiqatut Dianah

(150411100124)

Supervisor 1: Mula'ab S.Sc., M., Kom

Supervisor 2: Dr. Yeni Kustiyahningsih, S. Kom., M. Kom

ABSTARCT

Missing data sets are not the main problem for the substantive study in general, but it will cause serious problems if it is not resolved properly. There are many cases found in database of incomplete data that has half missing from the data set, in which it's very difficult to do data mining. Because only the complete data that can be analyzed. Thus, it was needed a more practical system which has the ability to fill in the missing values become possible values based on the available information in the data set. In addition, there are many methods to simulate the missing data. Weighted k-Nearest Neighbor (wkNN) method is a development of the k-Nearest Neighbor method (kNN) and It has the same principles as the k-Nearest Neighbor (kNN) method i.e. find the closest distance between the data to be tested with a number of k neighbors closest in the training data to determine the weight value. In this research. The data set was taken from a large amount of breast cancer data which was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. Based on the results of trials of 683 data, 9 attributes, 5 data which were eliminated sequentially and randomly, the results was obtained by using k nearest neighbor, that is k (3,5,7,9,11) and the best results are found in $k = 11$, because the greater of k which was used, the better the SSE value was obtained.

Keywords: Missing Data, wkNN, Nearest Neighbors, SSE